DOCUMENT RESUME

ED 275 166 FL 016 070

AUTHOR Berkoff, Nelson A.
TITLE Self-Assessment in English as a Foreign Language: Some Preliminary Ideas.
PUB DATE 84
NOTE 12p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; College Students; Comparative Analysis; Diagnostic Tests; *English (Second Language); Foreign Countries; Higher Education; *Language Tests; Rating Scales; *Self Evaluation (Individuals); *Student Placement; *Test Construction; Testing Problems; Test Reliability; *Test Use; Test Validity
IDENTIFIERS Israel

ABSTRACT

This study compares a self-assessment instrument and achievement test containing the same questions and administered at different levels of proficiency in order to investigate the reliability of second language self-assessment for student placement, diagnostic purposes, and achievement testing. Subjects were English as a Foreign Language students at Hebrew University in Jerusalem and students preparing to be English teachers at Teachers' Seminary. The instruments assessed vocabulary, grammar, listening comprehension, speaking ability, reading comprehension, and writing skills. Results suggest that (1) this form of test is unsatisfactory for foreign language placement of new students; (2) the most reliable self-assessment is found among students who had been studying for some time and actively use their skills; (3) self-assessment has great diagnostic value and can be used effectively as a formative test, during a language course; (4) students exposed to self-assessment become more conscious of their language strategies and thus better at self-assessment; and (5) this test type could be a valuable educational and psychological adjunct to teaching but is not yet proven reliable for second language achievement testing. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275166

SELF-ASSESSMENT IN ENGLISH AS A FOREIGN LANGUAGE: SOME PRELIMINARY IDEAS

NELSON A. BERKOFF
(THE HEBREW UNIVERSITY, JERUSALEM)

FL 016 070

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

P. Culhane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SELF-ASSESSMENT IN ENGLISH AS A FOREIGN LANGUAGE: SOME PRELIMINARY IDEAS

Nelson A. Berkoff (The Hebrew University, Jerusalem)

1 Some Preliminary Ideas

We spend a great deal of time preparing, marking and going through tests, but the one person who should be the most interested in the whole process - the student - is rarely asked to evaluate himself or even to have a say in the construction of tests. If we consider a normal teaching situation there are various points of time when a good, reliable, valid test is required - at the beginning and end of the course and at certain intervals along the way, when both students and teacher are interested in finding out what achievement or progress has taken place.

Consequently, I decided to investigate the possibility of constructing Self-Assessment Tests which might be suitable for all or any of these various points of time and, in particular, I had in mind the possibility of a reliable Self-Assessment Test that could be used as a Placement Test. It could be given to a student when he registers and, as a result of his own assessment, he could be allotted to the appropriate course. No time would be wasted in marking the test, and the student would have the feeling that he himself had decided at what level to study. Since this is not a do-or-die situation, the student would not be inclined to exaggeration in his self-assessment, for by doing so he would find himself in a class above his level and he would be the one to suffer. On the other hand, coming to a Self-Assessment Test "cold", (as is the case with our students who come to the university after three years of army service), the student might not be able to judge himself accurately.

In addition to its use as a Placement Test, a Self-Assessment Test would appear to have considerable diagnostic potential and, finally, once a reliable format and technique have been established it might be possible to use such a test as part of a Final Examination.

2 What Kind of Test?

The first problem was - what kind of test? Does one need a special kind of test or would a routine classroom test be suitable? A fairly obvious method is to ask the students - after they have finished a classroom test - to estimate the mark they think they should get on it; the test is then marked, returned to the students and the two marks compared. I presented a paper at the AILA Conference in Stuttgart in 1975 on some results of this method, but the whole operation is a fairly barren one, since very little can be learnt from it.

Another way is to use an X-point scale. In a Standardized Form

Method (as suggested by Oskarsson 1978), the student is asked to assess himself as follows: For Reading, "Imagine that someone with no knowledge of English comes to you with an English daily paper (e.g. The Times, The Guardian) and asks you to give as detailed an account as possible of what is on the front page. You have plenty of time but must not use a dictionary. How well would you manage the task?" (p. 40) The student is told to rate himself on an 11-point scale ranging from zero "I would not be able to say anything" to 10 - "I would be able to give a detailed account of everything on the page". This is extremely subjective and difficult. Another version suggested by Oskarsson is that of the "descriptive rating scale" in which each rank is spelled out; for example, the second highest rank in reading is "I understand everything or nearly everything written in the language within non-specialized fields. There may be words I do not understand in difficult texts"; and the one above the lowest is: "I understand the main point of a simple text and simple written directions for familiar things" (Oskarsson 1978:43). This is more satisfactory, but is still vague - what, for example, is a simple text?

I have recently been impressed by a method being used in Sweden today to test immigrant workers' knowledge of Swedish as a second language, in order to place them in an appropriate course (Von Elek 1982). This has two innovations: i) Students are not asked to carry out any language tasks, but they are asked whether they <u>could</u> carry out certain language tasks. For example:

a. <u>Vocabulary</u>. "Can you fill in the gap in the following sentence?"

   What _____ is that dress? It is red.

b. <u>Reading Comprehension</u>. "Do you understand this sentence?"

   Due to our well-organized pre- and post-natal care, the infant mortality is extremely low.

ii) The second innovation is that the students can give three possible answers:

1. Yes
2. I think so
3. No.

Von Elek found that the introduction of "I think so" was psychologically well-motivated, since it tended to prevent wild guessing and thus to make the "Yes" answer more reliable. In his research, the teacher went through parts of the test with each student. His provisional results showed that the students' self-judgements were highly reliable as far as the "Yes" results were concerned, the only major differences being with the grammar sub-section.

Although I did not feel so confident as Von Elek about the reliability of such a test as a Placement Test, it seemed to me to have great value for use <u>during</u> a language teaching course for three reasons:

1 <u>Educational</u>. Differences between students' self-assessment and

their actual achievement on the same questions should provide an opportunity for discussion between teacher and student as to the cause of such differences. It could result from lack of validity of the test, or from the student's unwillingness to face up to his weaknesses, or to his modesty or other factors which might not be revealed under normal circumstances. Consequently, I decided that immediately after the self-assessment test the students would be given a test with similar questions, but this time they would write in their answers and the two scores would be compared.

2 Diagnostic. Using my method makes it possible to pinpoint areas of weakness and/or uncertainty on the part of the student, and perhaps areas of weakness in the teaching. Once such areas have been discovered, ways of improving the teaching can be worked out. This, in fact, seems to be an improved form of Error Analysis.

3 Psychological. After being exposed to one or two self-assessment tests, the student will not only begin to have a much clearer idea of areas in which he is strong and those in which he is weak, but will also gain confidence as a result of this self-knowledge. This is something which is too often ignored.

Consequently, I have constructed self-assessment tests and achievement tests containing the same questions at different levels of difficulty, and used them with classes at different levels of proficiency. The aim was to investigate the reliability of such tests as Placement Tests, their potential value as diagnostic tests and the possibility of using them as part of an Achievement Test.

3 The Experiments

3.1 The Subjects

a. Students at the Hebrew University take a compulsory course in EFL with the main emphasis on Reading Comprehension. They study English for one or two years depending on their proficiency in English, and those studying for one year are divided into three levels. They may also study in an intensive summer course. Thus, we have four levels which I am calling Levels I, II, III & IV going from lowest to highest. Level I could be called Intermediate English and some of the material they study is edited and slightly simplified. Levels II, III & IV read only unsimplified material - Level II could be called Low Advanced, III Mid-Advanced and IV Advanced. It was expected that the differences in mean scores between levels on the same test would be about 10%. Students of Levels I, II & III have been tested - these are Groups P,Q & R.

b. Another group of students tested were prospective English teachers studying at a Teachers' Seminary. It was expected that they would be approximately at Level III - this is Group S.

3.2 The Tests

Von Elek had six sub-sections, testing Vocabulary, Grammar, Listening, Comprehension, Speaking Ability, Reading Comprehension and

Writing Ability. In our university course, although the student is generally expected to answer in English, little emphasis is put on this skill, which is not taught, and it was decided not to test it. On the other hand, students are listening to English all the time, and so Listening Comprehension was tested. In the pilot study Writing Ability was tested, although it is not taught; however, it turned out to be irrelevant and was dropped. Thus, the final tests which will be reported on, Tests B and C, consisted of:

a. Listening Comprehension - two short texts, read out by the classroom teacher, with five questions on each - 10 items.

b. Vocabulary - opposites and fill-ins - 15 items.

c. Grammar - Fill-ins and correcting incorrect sentences - 10 items.

d. Reading Comprehension - a short passage with five questions and a longer one with 10 - 15 items.

It was planned that Tests B & C would be approximately suitable for Levels II & III, and that the average Level II student achieve about 60% on Test B. Here I set my sights too high. Level A & D tests are now being constructed.

3.3 Administration of the Tests

In the self-assessment tests the questions were phrased in the form of: "Do you understand...?" or "Do you know...?" or "Can you supply...?" and so on. The students marked their answers in one of the three columns on an Answer Sheet: "Yes"; "I think so"; "No". Only the answers in the "Yes" column were scored as correct answers.

The achievement test consisted of exactly the same questions, but here the instructions were to "Write the opposite of the word underlined" or "Fill in the gap" or "Rewrite the sentence in a way that you think is correct".

3.4 Time Limit

It was decided not to limit the students for time. It was estimated that about 30 minutes would be required. Some finished in less than that time, but no one needed more than about 35 minutes on the self-assessment. On the achievement tests the students required about five minutes more. Once the tests become more or less standardized, it might be necessary to fix a time limit.

4 Results

4.1 Results on Test C

Table 1 shows the results of Test C, given to students who were expected to be at Level III. Group S were First Year Seminary students on their first day of study, and Group R were First Year University students on the second lesson of their course. For them the tests could be considered as Placement Tests.

Table 1

Results on Test C

	N	Self-Assess. Mean	S.D.	Achievement Mean	S.D.	Difference Overestimate (+) Underestimate(-)
Group S (III)	14	29.8	6.6	21.1	8.9	+8.7
Group R (III)	32	27.9	9.1	19.6	7.0	+8.3

4.2 Results on Test B

Table 2 shows the results of Test B, which was given to Groups P & Q whilst they were in the midst of an intensive summer course, and to Group S two days after they had taken Test C.

Table 2

Results on Test B

	N	Self-Assess. Mean	S.D.	Achievement Mean	S.D.	Difference Overestimate (+) Underestimate (-)
Group P (I)	18	25.3	6.4	20.5	8.3	+4.8
Group Q (II)	19	26.8	10.9	24.7	9.9	+2.1
Group S (III)	14	27.6	8.0	31.0	8.2	-3.4

4.3 Summary of Results on Tests B & C

For ease of comparison, Table 3 summarizes the results of Tests B & C, with the figures given in percentages.

Table 3

Summary of the results of Tests B & C. Figures in percentages.

Test B

	N	Self-Assess.	Achievement	Difference
Group P (I)	18	51%	41%	+10%
Group Q (II)	19	54%	49%	+ 5%
Group S (III)	14	55%	62%	-7%

Test C

	N	Self-Assess.	Achievement	Difference
Group S (III)	14	60%	42%	+18%
Group R (III)	32	56%	39%	+17%

5 Discussion

Since there are so many variables here, there is no point in looking for statistical significance, and so we will look at certain aspects of the data from a number of different angles:

5.1 Accuracy of Self-Assessment

a. At the beginning of a course

It was (mistakenly) thought that Level III students (Groups R & S) would find Test B too easy and would be insulted by being given such an "easy" test, and so they were given Test C. As can be seen from Tables 1 & 3 their achievements were low, about 40%, but what is interesting is that not only did both groups over-assess themselves but that their over-assessments are remarkably similar: 17% and 18%. Two days later Test C was returned to Group S, the results were discussed and they were given Test B. Possibly the shock of Test C sobered them because, although their achievement was reasonable - 62% - this time they under-assessed. (Unfortunately, it was not possible to give Test B to Group R.) Admittedly, the tests had been given to Group S in the wrong order, but the results show that self-assessment tests at the beginning of a course are not likely to be reliable.

b. In the middle of a course

Groups P and Q overestimated themselves on Test B (Table 2) but not excessively. However, on the skill that they had been mainly

taught, i.e. Reading Comprehension, the differences between self-assessment and achievement were negligible - it was +2% for Group P, -2% for Group Q. This certainly suggests that students are capable of evaluating themselves reasonably accurately on skills that they have been taught.

c. In general

The results (Table 3) indicate that if a test is "difficult" (achievement less than 50%), students tend to over-assess their proficiency. The only case of under-assessment was with Group S on Test B, which was a "reasonable" test for them, since their achievement was 62%. However, it may well be that the "shock" of Test C contributed to this under-assessment. Whatever the reason, we are in a vicious circle here - we want the students to assess themse-ves at the beginning of a course, but the most reliable test for them is one suitable to their level of proficiency and we will not know that till they have taken a test! This means that using this kind of self-assessment as a Placement Test in a foreign language (as opposed to a second language, as in Von Elek's research) is of doubtful value.

5.2 Relationship between positive and negative self-assessment and actual achievement

Tables 4 and 5 give the results of two questions arranged to show what percentage of students actually gave acceptable/unacceptable answers as compared with what they had marked or not marked on their Answer Sheet.

Question 84 was in the Grammar Section of Test B under the rubric: "Can you see what is wrong with the sentence and correct the error?" Question 84 was: "The news about it everyone was talking turned out to be false."

Question 143 was in the Reading Comprehension Section of Test C. The reading passage contained the following sentence: "There are theorists today who, proceeding on the assumption that all children are naturally good, kind, intelligent, unselfish and discriminating, deprecate any discipline or direction from adults." Question 143 was: "Could you fill in the TWO missing words? "The argument of the'theorists' (line 5) are _____ the assumption that all children are naturally good." The phrase "based on", which was the expected answer, appeared in the next sentence.

Table 4

Results of answers to Question 84. Figures are in percentages.

	Gave an acceptable answer			Gave an unacceptable answer			Gave no written answer		
	P	Q	R	P	Q	R	P	Q	R
Said that they knew	-	16	36	39	-	7	5	-	-
Said that they did not know	5	5	36	28	47	14	22	31	7

Table 5

Results of answers to Question 143. Figures are in percentages.

	Gave an acceptable answer		Gave an unacceptable answer		Gave no written answer	
	R	S	R	S	R	S
Said that they knew	19	42	22	14	-	-
Said that they did not know	9	29	12	14	-	-
Marked nothing on the Answer Sheet	9	-	3	-	25	-

From Table 4 we see that Question 84 was too difficult for Group P (Level I) and not much easier for Group Q (Level II). However, in Group S (Level III), 72% of the students gave an acceptable answer (36 + 36), but only half of those stated beforehand that they knew the answer. Again in Table 5, we see that 71% (42 + 29) of Group S gave an acceptable answer and this time well over half of them stated beforehand that they knew the answer. Group R (also Level III) would appear to be much less confident of themselves than Group S, since 34% (9 + 25) of them wrote nothing on the Answer Sheet and 25% gave no written answer.

These figures call into question the reliability of the whole research. What is happening here is that the overall picture that we get from Table 3 is masked by opposing differences, since + and - differences cancel each other out; but we do see that Group Q (Level II, whose mean score on the Achievement Test was 49%) show quite a stable picture - only one student (5% figure) estimating incorrectly. What has not been taken into account is the tendency of some people to over- and others to under-estimate themselves, irrespective of their language proficiency. This aspect was not investigated, but if such a "correction" could be introduced (by giving the student the appropriate psychological tests) it might clarify the picture. However, it would also make the whole procedure much too cumbersome.

5.3 Diagnostic Use

This is perhaps the most interesting aspect of the whole research. The incorrect rewritings of Question 84 are too many to list - here are a few examples of rewrites by students who claimed that they did know how to correct the sentence:

The news about everyone was talking turned out to be false.

The news about everyone was false.

The news was talking about turned out to be false.

The news about it which everyone was talking about turned out to be false.

Here are some of the unaccepted answers for Question 143 by students who claimed they did know the correct answer: against, to justify, to prove, proving that, about, keeping, believe that.

This is wonderful raw material for the teacher!

6 Conclusions

1 The kind of Self-Assessment Test, as used here, would seem to be unsatisfactory as a foreign language Placement Test for new students, although a series of graded tests presented in order of difficulty might produce more reliable results.

2 The most reliable area of Self-Assessment was with classes that had been studying for some time and with a skill, in this case ReadingComprehension, that they had been mainly taught. This produces the obvious conclusion that students can best evaluate themselves in the skills they have been taught!

3 Tests of self-assessment have great diagnostic value and they can be used with excellent effect during the progress of a language teaching course. They help both the student and the teacher to understand better the processes and strategies that the student is using.

4 Students who have been exposed to Self-Assessment Tests become more conscious of their language strategies and thus better at self-

assessment. Consequently, it seems feasible to use such tests as part of a Final Examination at the end of a language course, provided the students have had practice in them.

5 Finally, the type of Self-Assessment Test described here could be a valuable educational and psychological adjunct to the teacher's teaching, but it still has to be proved that it is a reliable method of self-assessment in a foreign language.

Bibliography

Berkoff, N.A. 1978. "Discrepancies between Student and Teacher Estimates of Language Proficiency". In Nickel, G. (ed). 1978. Language Testing. Papers from the Fourth World Congress of AILA. Stuttgart: Hochschulverlag.

Oskarsson, O. 1978. Approaches to Self-Assessment in Foreign Language Learning. Oxford: Pergamon Press.

Von Elek, Tibor. 1982. Tests of Swedish as a Second Language: An Experiment in Self-Assessment. Goteborg University: Language Teaching Research Center.